(2)

AD-A260 838

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
NAVAL AIR STATION, PENSACOLA, FL 32508-5700

NAMRL-1378

PERFORMANCED-BASED TESTING AND SUCCESS IN NAVAL ADVANCED FLIGHT TRAINING

David J. Blower

Approved for public release; distribution unlimited.

Reviewed and approved 25 NOV 92

A. J. MATECZUN, CAPT, MC USN
Commanding Officer

The research reported in this paper was completed under the Naval Medical Research and Development Command Work unit number 63706N M0096.002-M00960.01.

The views expressed in this report are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Trade names of materials and/or products of commerical or non-government organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commerical materials and/or products.

Volunteer subjects were recruited, evaluated and employed in accordance with the procedures specified in Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines.

SUMMARY PAGE

THE PROBLEM

Roughly 5% of student naval aviators fail the advanced phase of flight training. At this stage of training, the Navy has spent between $300,000 and $1,000,000 per student. Any reduction in this attrition rate through prior screening would be of great economic benefit to the Navy. Computer-based performance tests developed at the Naval Aerospace Medical Research Laboratory (NAMRL) were assessed to determine whether they could augment the present medical screening standards and thereby help identify potential failures in advanced flight training.

FINDINGS

A weak statistical relationship exists between a dual-task performance test, accession source, college major, an aptitude test, and success in advanced flight training. Discriminant analysis was employed to find a linear composite score of these variables that could be used to classify a student as a probable pass or fail in advanced flight training. For example, the model presented in this report could reduce failures by 50% at the cost of rejecting roughly 20% of those students who eventually passed. A Bayesian analysis of the success rate parameter showed that this particular model did result in a significant improvement over the present selection system.

RECOMMENDATIONS

These data can be used to make cost-benefit tradeoffs for aviation selection policy making. The author recommends that the dual-task performance test and accompanying statistical model discussed in this report be considered for operational implementation as part of an improved medical selection process for potential Navy and Marine Corps aviators.

Acknowledgments

I would like to acknowledge the valuable contributions to this report made by my colleagues in the Aviation Selection Division: LCDR Daniel L. Dolgin, LT David R. Street, Dr. Tatree Nontasak, Mrs. Kathy Helton, and to Dr. Harold Delaney of the University of New Mexico.

INTRODUCTION

In a previous report [1] computer-based performance tests were evaluated on how well they could predict the success of student naval aviators in the primary phase of flight training. Thereafter, the progress of these student aviators was monitored throughout the flight training program through the receipt of criterion data on grades, success, pipeline, *et cetera*, from Chief, Naval Aviation Training (CNATRA) in Corpus Christi, Texas.

When the data received during the intermediate and advanced phases of flight training were subsequently analyzed, no significant results were found between our battery of tests and success in the intermediate phase of flight training. We did, however, uncover a significant association between a dual-task performance test and success in advanced flight training.

These results are presented using a different method, as compared to the first report, for quantifying the importance of the data. Specifically, a linear composite of the test battery scores is constructed by using discriminant analysis. The discriminant function scores are then used to classify the students as to success or failure in advanced flight training. The success rate using these tests is then compared to the success rate under the present system. The formal technique employed here for making this comparison is the traditional Bayesian approach of calculating posterior probability density functions for the success rate parameter. Finally, confidence intervals are used to show the quantifiable consequences for Navy planners if they should decide to use these tests for selection purposes.

A counter-argument based on the hypothetical inflexibility of the training command infrastructure due to fixed costs is presented in the Discussion. If such an argument has any merit, it tends to vitiate any optimism for implementing these tests for reasons due solely to the statistical results.

METHODS

SUBJECTS

Student naval aviators, preselected for naval aviation flight training on the basis of their performance on the current Navy and Marine Corps aviation selection tests and medical examinations, participated in the study. The actual number of subjects for specific cases are presented in the Results section. The subjects were informed that a) the investigation involved performing tasks in problem solving and perceptual and motor skills, b) their test performance would not affect their continuation in the program nor be entered into their permanent service records and, c) results would be used solely for the purpose of developing an improved aviation selection program for the Navy.

APPARATUS

All testing was conducted on Apple IIe microcomputers with Apple monochrome monitors (CRTs). Subjects used a numeric keypad to respond to discrete stimuli. All responses were recorded to millisecond accuracy. A Measurement Systems Incorporated (MSI-542) control stick was used for joystick and throttle control during the tracking tasks. Rudderpedal controls were measured using a variable resistor connected to a computer A/D channel. The joystick was mounted on the forward edge of the testing console at a centered position. The throttle was located on the left side of the testing booth. The rudderpedals were located so that the subjects could easily place both feet while sitting in the testing booth. Subjects operated the joystick with the right hand, the throttle with the left hand, and the rudderpedals with both feet.

PROCEDURE

All candidates were tested before entering flight training and after completing a 14-week basic military indoctrination program for AOC officers or a 6-week program for students already commissioned. All instructions were presented to the subjects on the CRT for each task individually. Test administrators intervened only to begin the computer program for each task and to answer questions posed by subjects. The test administration time of the battery ranged from 3.7 to 4.0 h. The order of the tasks and the stimuli within each test were identical for all subjects. Subjects received a 3-4 min rest period between tasks. All testing took place in an air-conditioned laboratory.

PREDICTOR TESTS

The entire complement of tests in the NAMRL computer-based performance test battery is described in [1]. We repeat the description of the Horizontal Tracking test, the Absolute Difference test, and the combined Absolute Difference-Horizontal Tracking test because these tests exhibited an association with success in advanced flight training. Damos and Gibb [2] explain the rationale for these tests and present some preliminary results comparing the performance of experienced fleet pilots to student naval aviators on these tests.

Several background variables were recorded for each subject. These included:

1. Initial aviation selection test scores, *i.e.* Academic Qualification Test (AQT) and Flight Aptitude Rating (FAR)

2. Previous civilian flight training

3. Age

4. Source of procurement

5. Gender

6. College major

Previous civilian flight training was treated as a continuous variable using the self-reported number of total flight hours a subject had logged. Total flight hours included both solo and dual pilot training hours.

Source of procurement included six distinct groups:

1. Aviation Officer Candidate School (college graduates entering directly into the military)

2. Naval Academy graduates

3. Naval cadets (prior enlisted service or 60 college credits with numerous other criteria)

4. Marine Corps Officer Program

5. Naval Reserve Officer Training Corps

6. Other (direct procurement, Merchant Marine Academy, enlisted commissioning programs)

College major was classified into one of five general disciplines:

1. Engineering and math

2. Physical sciences (biology, geology, physics, etc.)

3. Business

4. Social sciences (psychology, sociology, history, etc.)

5. Physical education

HORIZONTAL TRACKING

In this test, the subject was required to learn a one-dimensional tracking task. To perform the test successfully, the subject had to anticipate the movement of a square on the computer screen and manipulate the joystick to counterbalance the movement in order to keep the square centered on a fixed central point on the screen. For example, if the square was moving off center to the right, the subject would move the joystick to the left in order to re-center the square. Specifically, the subject maintained a 0.6-cm square centered in a 9.75 by 1.25-cm rectangle by moving the joystick either left or right. The square was driven by a forcing function programmed into the computer. Each subject received ten 2-min trials separated by a 30-s rest. The dependent measure was RMS error. Total testing time was 25-min. This test was designed to measure compensatory tracking skills. A compensatory tracking task is defined as one in which only the difference (the error) between the command input and the system output is displayed. [3]

ABSOLUTE DIFFERENCE TASK

In this test, the subject was visually presented with a random number between 1 and 9 on the CRT. This number disappeared from the screen and was then followed immediately with another number. The subject was required to press the correct button on the numeric keypad to indicate the absolute difference between the number presently displayed on the CRT screen and the number shown on the screen in the previous trial. The numbers were presented in such a manner that only absolute differences from 1 to 4 were correct. When the subject made a response, a new number appeared on the CRT, and again the subject calculated the absolute difference between the number previously presented and the number currently displayed. The subject was instructed to use only his right hand in responding with the numeric keypad. Speed and accuracy of response were equally emphasized. This test was self-paced and consisted of fifteen 2-min trials. Total testing time was 30 min. The dependent measures consisted of number correct, number of errors, RT on correct responses, and RT on error responses. This task was essentially a measure of short-term memory, memory search, and encoding.

ABSOLUTE DIFFERENCE AND HORIZONTAL TRACKING

As part of the emphasis on time-sharing ability within the NAMRL battery, the horizontal tracking and absolute difference tasks were combined for a measure of dual-task performance. For this combination of tasks, the subject performed the horizontal tracking task and the absolute difference task simultaneously. The number for the absolute difference task was centered above the tracking task and touched the top of the tracking task. The subject controlled the tracking task with his right hand and, for the absolute difference task, pressed a number on the keypad with his left hand. The subject was instructed that the two tasks were equally important. The subject received five 2-min trials separated by rest periods of 30-60 s. Total testing time for the dual-task combination was 15 min. Figure 1 depicts this dual-task test as it appeared to the subject on the computer screen.

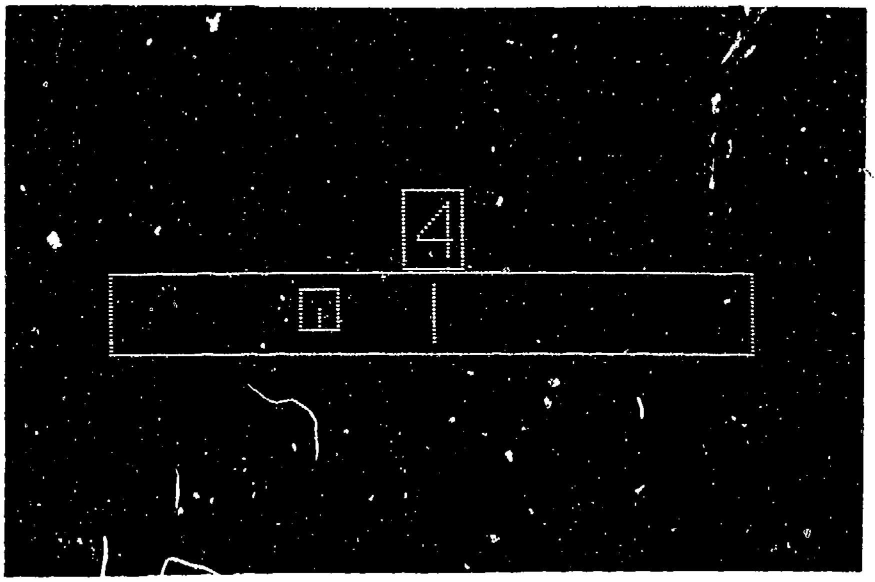

Figure 1. The one-dimensional tracking test and absolute difference test combined to form a dual-task test.

RESULTS

DATA BASE AND PREPROCESSING OF VARIABLES

The original data base consisted of 144 variables on 1293 student naval aviators (SNAs). Since the previously cited report on primary flight training, additional data from CNATRA have been received on the intermediate and advanced phases of flight training. When these data were added to the data base, there was pass or fail information for the advanced phase of flight training on 836 SNAs. There was a substantial amount of missing data for these 836 subjects because individual tests were introduced into the test battery at different points in time.

Appendix A contains descriptive data in the form of cross-tabulation tables for success in advanced flight training versus selected background variables. The frequency table for Training Success is presented first, followed by the cross-tabulation tables for Training

Success versus Gender, Accession Source, College Major, AQT, FAR, and Age. The frequencies for these variables, when the missing data are included, add up to 836.

In order to cope with this mass of data, we used hierarchical multiple regression to screen out promising variables for attention.[1] When the dichotomous variable of success in advanced flight training was employed as the criterion variable, only one set of variables reached statistical significance. After the variables of age, previous flight hours, sex, accession group, college major, AQT, and FAR had been entered in the regression equation, only the eight variables from the AD, HT, and dual-task ADHT reached significance as measured by the increase in R^2. [4]

At this point, the number of variables had been pared to 15. We subjected these 15 variables to further analysis. First, we desired to construct a composite score based on the eight ADHT variables. One feasible composite score was arrived at by utilizing only the dual-task ADHT test in the following manner:

$$ADHTCS = .20 * ZADHT6 - .50 * ZADHT5 - .10 * ZADHT7 - .20 * ZADHT8$$

This composite score consisted only of the four variables from the dual-task ADHT test, i.e. when subjects were performing the absolute difference task and the horizontal tracking task at the same time. Because the subjects had been instructed to pay equal attention to both parts of the test, the composite score weighted the one variable concerned with tracking error (ZADHT5) equally with the three variables concerned with absolute difference performance (ZADHT6, ZADHT7, and ZADHT8). These last three variables together were given a weight equal to 0.5 with the further breakdown that mean number of errors (ZADHT7) was given only half the weight of mean number correct (ZADHT6) and mean correct RT (ZADHT8). All four scores were first converted to z-scores before the composite score was formed. Any constituent z-score making up the ADHTCS greater than 4 in absolute value was rejected as an outlier.

Success in advanced flight training was coded as a "1" and failure as a "0." Performance on AD number correct (ZADHT6) was the only variable *positively* correlated with the success code. The other three variables, tracking error (ZADHT5), AD number of errors (ZADHT7), and mean correct RT (ZADHT8), were negatively correlated with the success code. This pattern of correlations explains the signs of the weighting coefficients in the composite score so that ADHTCS as constructed was positively correlated with the success code.

To illustrate, consider the computation of the ADHT composite score for a hypothetical SNA who did much better than average on the tracking portion of the test, but performed somewhat poorer on the absolute difference segment. His z-score for tracking (ZADHT5) was −2. (Remember that a lower z-score for tracking error represents better performance). On the other hand, his z-score for mean number correct on the absolute

[1] This and other related work was carried out by Dr. Harold Delaney of the University of New Mexico while he held an ASEE Summer Faculty appointment at NAMRL in 1991.

difference segment (ZADHT6) was -1, his z-score for mean number of errors (ZADHT7) was $+1.3$ (a higher z-score represents poorer performance), and his mean correct RT (ZADHT8) was $+1$ (again, a higher z-score represents poorer performance). When these values are inserted into the composite score equation, ADHTCS equals $+0.47$.

Furthermore, of the background variables and aptitude tests, only accession group, college major, and AQT played a significant role. The results reported here concern these three variables and the ADHT composite score. Since accession group and college major were both coded with two dummy variables, the analysis will deal with six variables in all.

DISCRIMINANT ANALYSIS

A discriminant analysis (DA) was conducted on these variables with the SPSS/PC+ statistical package [5]. Only cases with complete data on all variables were subjected to analysis. Of 451 such cases, 432 students passed advanced flight training (PASS), and 19 students failed advanced training (FAIL).

There were 438 males and 13 females constituting the 451 SNAs with complete data. Their age ranged from 20 to 29 with a mean age of 23.04 and a standard deviation of 1.43.

The relevant summary statistics from the DA were as follows. The canonical correlation equaled .1712, indicating a slight influence of these six variables on success in advanced flight training. A $\chi^2 = 13.274$ suggests that the means of the discriminant function scores for the PASS group and the FAIL group were significantly different from zero ($p < .039$, $6df$). The covariance matrices for the two groups did not depart from equality by Box's M test.

Table 1 illustrates in detail the calculation of the discriminant score for one SNA. The six variables plus a constant are shown in the first column. d_1 and d_2 are dummy variables coding accession group, while d_3 and d_4 are dummy variables encoding college major. Table 2 shows the relationship between these variables and their dummy variable coding. The second column of Table 1 gives the *unstandardized* discrimination function coefficients as computed by the program. The third column shows the values for a hypothetical subject on the six variables. This subject had an AQT stanine score of 7, (on a scale of 1–9), and was a Naval Academy accession ($d_1 = 0$ $d_2 = 0$) with an engineering/mathematics degree ($d_3 = 1$ $d_4 = 0$). We shall use the previously computed ADHT composite score of .47 for this SNA.

The mean discriminant score for the entire FAIL group in the sample was .827, and the mean discriminant score for the entire PASS group was $-.036$. The mathematical details of how the students are classified into a predicted PASS or FAIL group based on the discriminant score are given in appendix B. For the hypothetical SNA given in

Table 1. Unstandardized discrimination function coefficients and values for six variables plus a constant needed to compute a discriminant function score for a hypothetical subject.

Variable	*Coefficient*	*Value*	*Multiplication (x_i)*
AQT	.263	7	1.843
d_1	.586	0	0.000
d_2	−.238	0	0.000
d_3	1.474	1	1.474
d_4	.735	0	0.000
ADHTCS	−1.324	0.47	−0.622
Constant	−2.531	1	−2.531
		Discriminant Function Score	$\sum_{i=1}^{7} x_i = .164$

Table 2. The dummy coding for the accession source and college major variables.

Accession Source	d_1	d_2
Naval Academy	0	0
AOC/OCS	1	0
ROTC/Marines	0	1
College Major	d_3	d_4
General Science	0	0
Engineering/Math	1	0
Liberal Arts	0	1

Table 3. The classification matrix arising from the discriminant analysis of 451 SNAs showing the predicted passes and fails of the model compared to the actual data.

		PREDICTED Pass	PREDICTED Fail	
ACTUAL	Pass	339	93	432
	Fail	9	10	19
		348	103	451

Table 1, the calculations in appendix B assign a probability of passing advanced flight training of 63%. The SNA would, therefore, be assigned to the PASS group.

Table 3 shows the classification matrix that is produced when the DA program classifies each of the 451 known pass or fails into a predicted pass or fail. The prior probabilities parameter in the DA program was adjusted so that close to a 50% attrition rate would be achieved. From Table 3, observe that 10 of 19 students have been correctly predicted into the fail category. The observed success rate of 97.41% (339/348), when using these variables to discriminate the PASS and FAIL groups, is achieved at the cost of rejecting 93 SNAs who otherwise would have passed.

Figure 2 shows the theoretical Gaussian distribution of the discriminant function scores. The means are set to the values as calculated by the program, $\mu_{PASS} = -.03637$ and $\mu_{FAIL} = .82694$, with $\sigma_{PASS} = \sigma_{FAIL} = 1$. The threshold discriminant score at .769 divides the two curves into four areas. The number of SNAs falling into each of these four areas when the threshold discriminant function score is used to classify SNAs is also shown. This is the same information presented in the classification matrix of Table 3.

Logistic regression is sometimes recommended as an alternative to discriminant analysis for finding the weights to attach to the variables that form the linear combination used subsequently to classify cases. The logistic model may be more appropriate when the explanatory variables, such as accession source and college major, are qualitative and obviously not distributed as multivariate normal.

Accordingly, a logistic regression was carried out using the same six explanatory variables as in the DA model and with training success as the dependent variable. The logistic regression parameters were estimated by maximum likelihood. These parame-

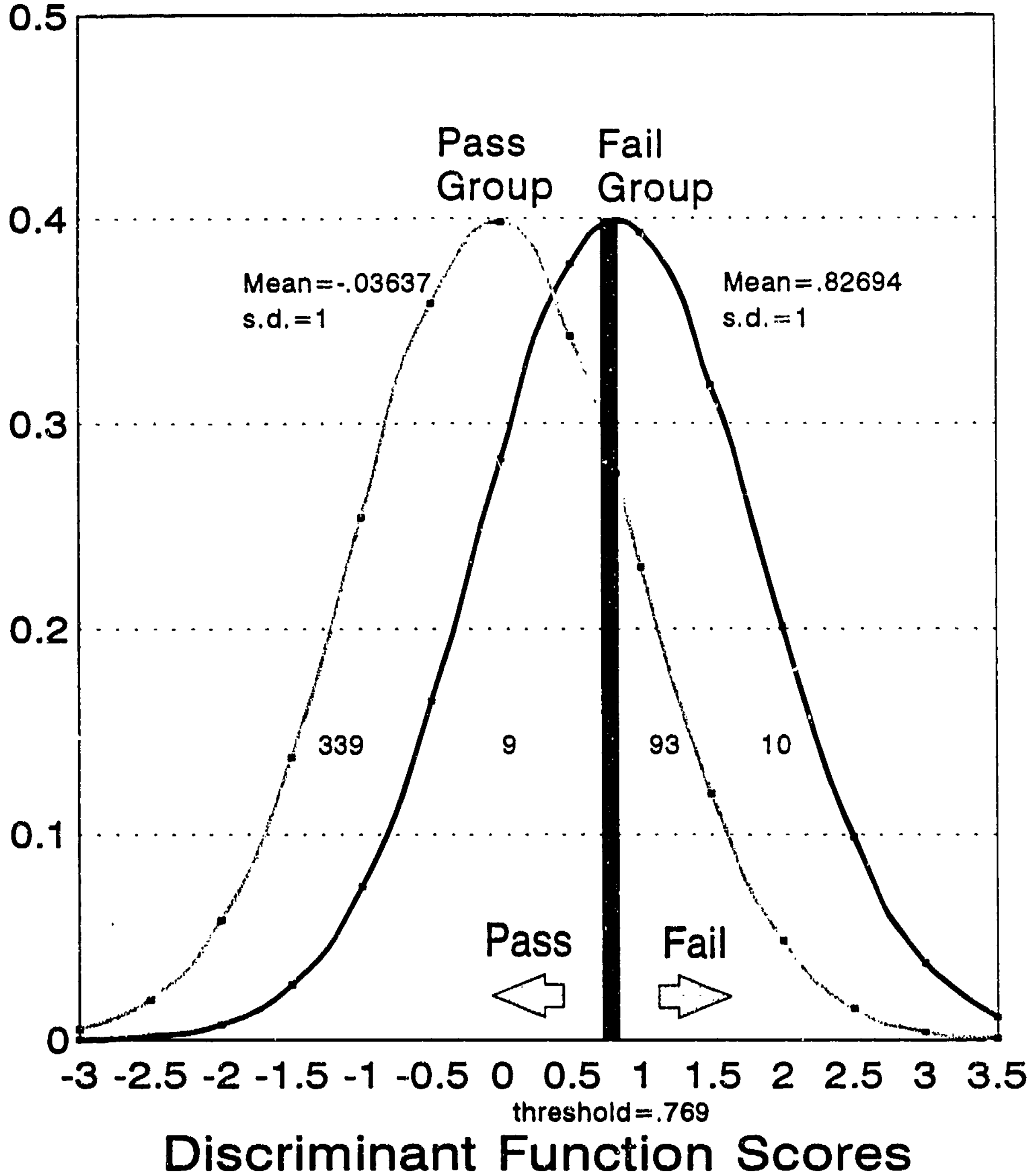

Figure 2. The theoretical Gaussian probability density functions of the discriminant function scores for the PASS and FAIL groups in advanced flight training.

Table 4. Logistic regression coefficients for the same six variables and hypothetical values used in the DA model. The logistic regression coefficients are meant to be compared with the DA coefficients shown in Table 1.

Variable	*Coefficient*	*Value*	*Multiplication*
AQT	−.231	7	−1.614
d_1	−.566	0	0.000
d_2	.246	0	0.000
d_3	−1.690	1	−1.690
d_4	−1.068	0	0.000
ADHTCS	1.071	0.47	.503
Constant	6.065	1	6.065
	Logistic regression score		$\sum_{i=1}^{7} x_i = 3.26$

ter estimates are presented in Table 4 where they may be compared to those found in Table 1. Except for a change in sign and a shift due to the constant, these coefficients are comparable to the coefficients found using the DA estimates. The same hypothetical SNA is used to compute a logistic regression score of 3.26 when using these coefficients.

The mean of the logistic regression scores was 3.522 for the PASS group and 2.745 for the FAIL group. The posterior probability of belonging to the PASS group for a given logistic score was calculated, as explained in appendix B, in the same fashion as for the DA model. The hypothetical SNA with the logistic regression score of 3.26 has a posterior probability of 60% of passing as compared to 63% for the DA model. This SNA would still be classified as belonging to the PASS group.

The classification matrix shown in Table 5 results when these probabilities are used to classify all 451 SNAs. We can see that the classification based on the use of the logistic regression weights hardly differs from the outcome based on the DA weights.

BAYESIAN ANALYSIS

The success rate parameter, θ, in advanced flight training is now dealt with by a standard Bayesian analysis [6]. Two situations are of interest in this report. These two situations involve the comparison of the posterior probability density functions (PPDF) for θ when the current system selects students and the PPDF for θ when the discriminant

Table 5. The classification matrix arising from the logistic regression classification of 451 SNAs. This table can be compared to Table 3 which presents the DA classifications.

		PREDICTED		
		Pass	Fail	
ACTUAL	Pass	344	88	432
	Fail	10	9	19
		354	97	451

analysis model is used to select students. The success rate parameter θ is considered to be a continuous parameter with $0 < \theta < 1$. The likelihood of success in advanced flight training is assumed to follow a binomial distribution. The merit of using the DA model discussed in this report then hinges on the characteristics of its PPDF as compared to the PPDF of the present system. The numerical details of this analysis are contained in appendix C.

The prior probability density function shown in Fig. 3 is based on the calculations given in Appendix C . This prior was selected to represent our state of knowledge based on historical records for success rate in advanced flight training.

Of the 836 SNAs in the data base, 788 passed advanced flight training, and 48 failed. The PPDF constructed using this information on success rate is also shown in Fig. 3 as the solid curve. The mean of this PPDF is .9434, the mode is .9443, and the standard deviation is .0076.

The effect from the additional information contained in the DA model on the PPDF is illustrated in Fig. 3 as the dotted curve. Of the 836 SNAs in the data base, 451 SNAs had data on the ADHT test. Referring back to the classification matrix in Table 3, focus attention on the column the model predicted to pass. Of these 348 students, 339 students passed and 9 failed. The mean of the PPDF using this information is .9688, the mode is .9709, and the standard deviation is .0082. Since the DA model is based on a smaller sample size than the present selection system, which makes use of the entire data base, it is understandable that the standard deviation for the PPDF based on the DA model is correspondingly larger.

Therefore, Fig. 3 shows the prior and the two PPDFs together on the same graph

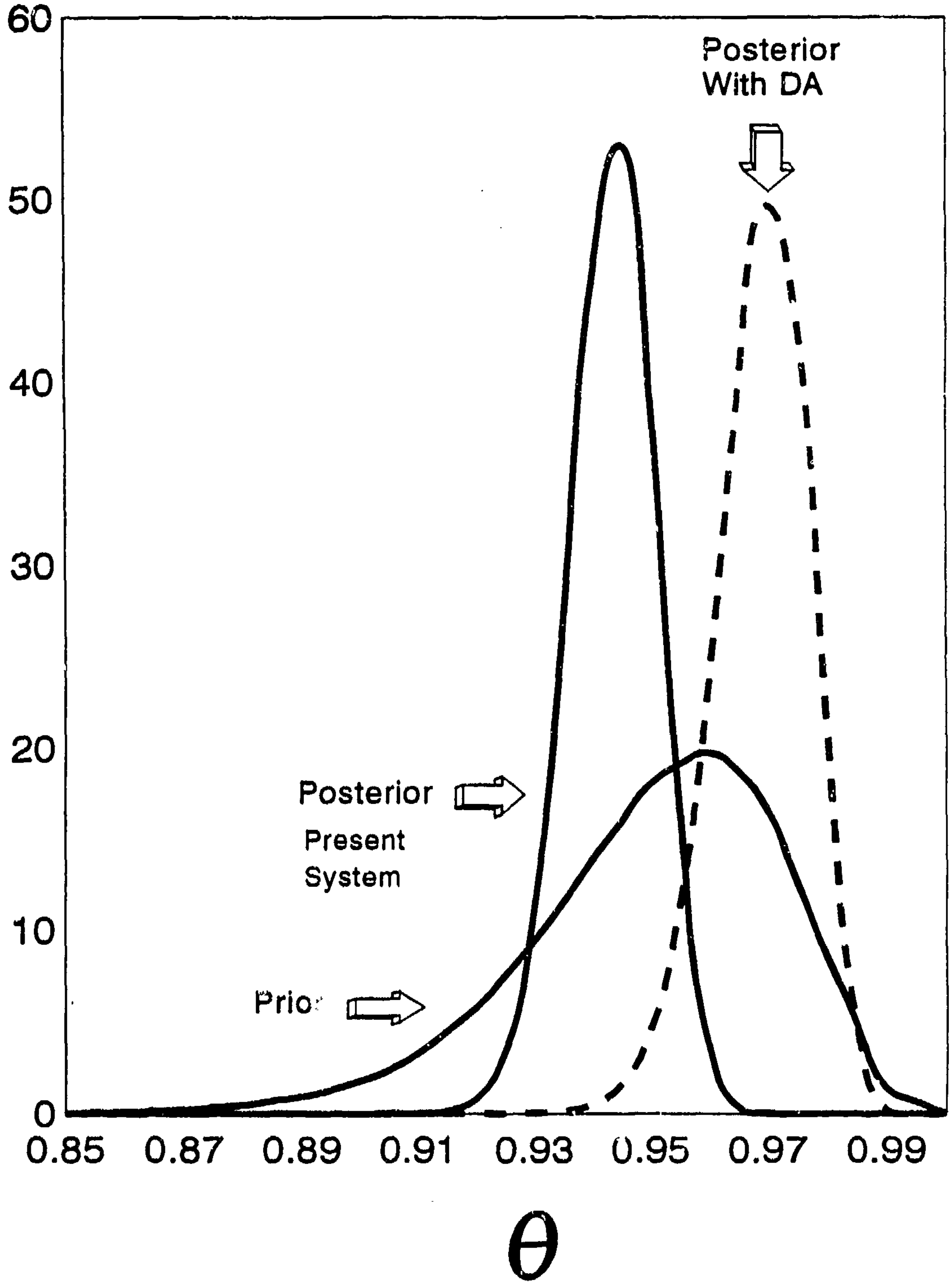

Figure 3. The prior and posterior probability density functions for the success rate parameter θ in advanced flight training. The solid curve represents information using the present selection system and the dotted curve represents information from the DA model.

Table 6. Various Bayesian confidence intervals for the posterior probability density functions for the present selection system and one using the DA model.

	Present System	DA Model
Mean	94.34%	96.88%
90% CI	93.19 to 95.68%	95.72 to 98.46%
95% CI	92.90 to 95.86%	95.40 to 98.78%
99% CI	92.42 to 96.44%	94.68 to 99.60%

for an overall visual comparison. The PPDF using information from the ADHT test is shifted to the right, illustrating the improvement in the success rate as compared to the current selection system.

Table 6 presents a final characterization of the two systems by calculating the 90, 95, and 99% Bayesian confidence intervals[2] for the respective PPDFs. There is no overlap in the 90% confidence intervals for the two systems. The confidence intervals for the DA model are wider than the corresponding confidence intervals for the present system because of the difference in sample size.

DISCUSSION

The coefficients of the variables in the DA model deserve some further comments in light of the cross-tabulation tables presented in appendix A. The failure rates given there are based on the larger sample size available for the background variables as opposed to the smaller sample size when the ADHT was used in the DA model.

The positive coefficient for the AQT aptitude test means that a *better* score is linked to a *lower* probability of success in advanced flight training. (Remember that higher discriminant function scores result in lower probability of success and vice versa). Although it is contrary to intuition that SNAs with superior verbal and visuospatial skills have a reduced chance of success in flight training, there is some evidence in the literature for this phenomenon. [7]

[2]The terms credibility interval and highest posterior density (HPD) are also used in the Bayesian literature.

The cross-tabulation for AQT supports this finding, especially in the range of scores from 4 to 8 where the majority of the data are found. The probability of failure increases from near 4% with an AQT score of 4 to over 10% with an AQT score of 8. It remains somewhat disconcerting that a test the Navy uses to select individuals for training eventuates to those having higher scores doing less well in advanced flight training.

In order of decreasing probability of success, the ranking of the accession source variable (d_1 and d_2) is,

1. ROTC/Marines
2. Naval Academy
3. AOC/OCS

The AOC/OCS accession source has traditionally done less well in all phases of flight training, so their ranking here is not surprising. Perhaps slightly more surprising, though, is the second place finish of the Naval Academy graduates. In our previous analysis of success in primary flight training, the Naval Academy graduates had outperformed the other two accesssion sources.

The cross-tabulation data show essentially no difference between the Naval Academy graduates and the ROTC/Marines. The failure rate for these two groups hovers around 4.5%. The failure rate of the AOC/OCS group is clearly higher at around 7%.

College major, (d_3 and d_4), had the ranking of

1. General Science
2. Liberal Arts/Other
3. Engineering/Math

This seems anomalous as one would expect the engineering/math majors to be the more successful group. It may be that excessively analytical tendencies in some few individuals with engineering/math backgrounds are antithetical to "seat-of-the-pants" flying [7], although this remains the merest sort of speculation.

The cross-tabulation data support the superiority of the general science majors vis-a-vis the other two college major groupings with a failure rate of about 2.5% for general science majors versus failures rates over 6% for the others. These data do not, however, support any distinction between the engineering/math majors and the liberal arts/others major as the DA model does.

Gender and Age were not included in the DA model, but, interestingly, note from the cross-tabulation tables that the failure rate for women is comparable to that for the men. Of course, the overall number of women in the data base is extremely small, but still accurately reflects the relative number of women in the training program. The Age cross-tabulation table seems to suggest that increasing age results in poorer training performance.

Returning to the DA model, the dual-task composite variable (ADHTCS) with its negative coefficient was in the expected direction with higher scores leading to increased probability of success. Although a strong relationship does not exist between the ADHT composite score and success in advanced flight training, the attrition rate *can* be reduced if the resulting cost of false rejections is acceptable.

For example, the DA model presented in this report could achieve a 50% reduction in attrition rate at the cost of roughly a 20% false rejection rate. That is, about 20% of the SNAs would attain a discriminant function score that would categorize them as failing when, in fact, they would have passed. The classification matrix of Table 3 illustrates this fact. Other desired attrition rates would entail different false rejection rates. These inferences about a higher success rate are related to the particular variables chosen and the specific linear composite of these variables as formed by the DA program, as well as the prior probabilities and the costs associated with correct and incorrect decisions.

One rather naive approach to calculating cost savings is to multiply the expected number of SNAs not needed to be trained by the cost per student for the appropriate pipeline. In this numerical example we shall assume that 1000 aviators enter advanced flight training. Under the present selection system we would expect that 5%, or 50 SNAs, would attrite from the program. If one were to adopt the threshold cut-off score from the DA model resulting in a 50% reduction in attrition rate, only 2.5%, or 25 SNAs, would attrite. Thus, 25 students would not need to be trained, resulting in a cost savings. Table 7 presents some not unrealistic training costs per student for the situation where the 25 students are evenly spread out over the three pipelines. Even considering the cost of false rejections, some significant monetary gain is achievable.

Now, a straightforward calculation similar to that above, which might save $13.6M per year seems to support the argument for implementation of the additional tests. The actual situation, however, is probably more subtle than this naive calculation.

The following question can be posed: *"What changes to the training infrastructure would have to be made in order to realize these projected savings?"* A little thought would seem to indicate that the following factors would be among the major components involved in training costs.

1. The number of aircraft needed to carry out the training curriculum.

2. The number of flight instructors to fly these aircraft.

Table 7. One way of calculating cost savings in advanced flight training when attrition is reduced from 5% to 2.5%. The calculation is based on 1000 aviators.

Pipeline	Number	Savings	Total
HELO	8	$300,000	$2.4M
PROP	8	$500,000	$4.0M
JET	9	$800,000	$7.2M
	25		$13.6M

3. The number of hours flown during the training curriculum to reach some desired level of proficiency.
4. The cost of fuel.
5. The cost of ground personnel and other support personnel.
6. The cost of maintaining the training aircraft.
7. The cost of simulators and simulator support.
8. Billeting the students while they are in training.

The follow-on question then becomes *"How elastic are these various factors in response to reduced attritions in order for savings to be achieved?"* If, in fact, they are rather inelastic, then the savings postulated above in Table 7 would be difficult to realize.

This is simply an inquiry about the inherent rigidity in the training infrastructure. If the Navy has already bought the planes for a planned attrition rate, set up the billets for flight instructors and ground personnel, and derived a curriculum that says students shall fly these number of hops and these number of hours, then it would be difficult to adjust these factors because of small changes in the attrition rate. Can the Navy actually buy 2 fewer planes, reduce 5 instructor billets, and fly 30 hours less in the advanced curriculum simply because they have 25 fewer pilots to train? We also have to remember that these advanced training costs are spread out over three different pipelines and conducted at different training bases where presumbably the costs are again relatively fixed.

Perhaps, an easier way of putting this into perspective is to look at the numbers involved in advanced flight training. If the Navy has a training infrastructure set up to train 1000 students, can it really trim that infrastructure because it now has "only" 975 students?

It doesn't seem logical that the large savings as indicated by the calculations made above are completely warranted if the rigidity of the training infrastructure is a reality. It would be highly desirable to submit the scenario outlined above to an expert in the economic aspects of training for a critical analysis. The decision to implement additional selection tests clearly depends upon such an analysis.

A second, more subtle, reservation about realizing savings due to slight reductions in attrition rates concerns the "mental set" of the flight instructors. If some sort of "corporate memory" or "corporate culture" exists in the training commands that dictates adherence to attriting 5% of the students, then will or can the instructors adjust to attriting only 2.5% of the students? Is each individual flight instructor able to fine tune his assessment of flight proficiency so that a slightly smaller number of students is attrited over all pipelines and over all training commands? In addition, flight instructors might not feel like they are doing their job if they don't fail some minimum number of students. So the net effect in the end might be that, although additional tests are effective and able to predict some of the students who would have failed in advanced, the flight instructors will ratchet up their grading so that the "expected" number of failures occur anyway, thus offsetting the effect of the additional tests.

CONCLUSION

A weak statistical relationship exists between a dual-task performance test, accession source, college major, an aptitude test, and success in advanced flight training. For practical planning purposes, it is assumed that the decision to implement additional tests would be based on meaningful increases in the success rate. Information about success rate from our data is neatly encapsulated in the posterior probability density functions shown in Fig. 3. These PPDFs show that the success rate, using the variables and model mentioned above, is higher than the success rate of the present selection system, which does not take these variables into account. More precisely, the 90% Bayesian confidence intervals for these two density functions do not overlap, leading one to the belief that success rates for these two selection systems really are different. This higher success rate entails a cost in terms of rejecting some candidates who would have been successful. Whether the statistical differences in success rates can be translated into actual dollar savings at the training command infrastructure level remains an open question.

References

[1] Blower, D. J. and Dolgin D. L., *An Evaluation of Performance-Based Tests Designed to Improve Naval Aviation Selection*, NAMRL 1363, Naval Aerospace Medical Research Laboratory, Pensacola, FL, August 1991.

[2] Damos, D. L. and Gibb G. D., *Development of a Computer-Based Naval Aviation Selection Test Battery*, NAMRL 1319, Naval Aerospace Medical Research Laboratory, Pensacola, FL, August 1986.

[3] Damos, D. L. Ed., *Multiple Task Performance*, Taylor & Francis Ltd., London, UK, 1991.

[4] Delaney, H. D., *Analysis of Performance-Based Tests for Advanced Flight Training*, Unpublished manuscript, ASEE Summer Faculty Program, Naval Aerospace Medical Research Laboratory, Pensacola, FL, July 1991.

[5] Norušis, M.J., *Advanced Statistics SPSS/PC+*, SPSS Inc., Chicago, IL, 1986.

[6] Press, S.J., *Bayesian Statistics: Principles, Models and Applications*, Wiley, New York, NY, 1989.

[7] Gordon, H.W. and Leighty, R., Importance of Specialized Cognitive Function in the Selection of Military Pilots. *Journal of Applied Psychology*, Vol.73, pp. 38-45, 1988.

[8] Tatsuoka, M.M., *Multivariate Analysis: Techniques for Educational and Psychological Research*, Wiley, New York, NY, 1971.

[9] Mathcad reference manual, Ver 3.1. Mathsoft, Inc., Cambridge, MA, 1988.

APPENDIX A

This appendix contains descriptive data on the backgound variables and training success in advanced flight training for the 836 SNAs in our data base. The tables, except for the first, are presented as cross-tabulation tables showing the breakdown of gender, accession source, college major, AQT, FAR, and age with training success. The failure percentage for each level of these variables is also presented. The total number, when the missing data are included, add up to 836 for each variable.

TRAINING SUCCESS		
Pass	788	94.26%
Fail	48	5.74%
Total	836	

GENDER				
Level	Pass	Fail	Total	% Fail
Male	770	47	817	5.75
Female	17	1	18	5.56
Missing			1	

ACCESSION SOURCE				
Level	Pass	Fail	Total	% Fail
Naval Academy	173	8	181	4.42
ROTC/Marines	229	11	240	4.58
AOC/OCS	381	29	410	7.07
Missing			5	

COLLEGE MAJOR				
Level	Pass	Fail	Total	% Fail
Engineering/Math	350	23	373	6.17
General Science	113	3	116	2.65
Liberal Arts/Other	320	22	342	6.43
Missing			5	

AQT				
Level	Pass	Fail	Total	% Fail
2	1	0	1	0.00
3	18	4	22	22.22
4	94	4	98	4.08
5	262	12	274	4.38
6	206	11	217	5.07
7	146	12	158	7.59
8	44	5	49	10.20
9	17	0	17	0.00
Missing			0	

FAR				
Level	Pass	Fail	Total	% Fail
3	12	1	13	7.69
4	23	3	26	11.54
5	77	0	77	0.00
6	171	10	181	5.52
7	144	6	150	4.00
8	140	9	149	6.04
9	221	19	240	7.92
Missing			0	

AGE				
Level	Pass	Fail	Total	% Fail
20	3	1	4	25.00
21	40	1	41	2.44
22	303	17	320	5.31
23	201	7	208	3.37
24	114	9	123	7.32
25	61	7	68	10.29
26	46	4	50	8.00
27	11	1	12	8.33
28	4	1	5	20.00
29	2	0	2	0.00
30	1	0	1	0.00
Missing			2	

APPENDIX B

The SPSS/PC+ Advanced Statistics Manual explains in general terms how cases are classified by the discriminant analysis program. "A case is classified, based on its discriminant score, in the group for which the posterior probability is largest." [5]. Bayes's Theorem is used to calculate the posterior probability

$$P(G_i|D) = \frac{P(D|G_i) * P(G_i)}{\sum_{i=1}^{S} P(D|G_i) * P(G_i)}$$

where $P(G_i|D)$ is the posterior probability for Group i, $P(D|G_i)$ is the likelihood of the data given Group i, and $P(G_i)$ is the prior probability of Group i.

For the two-group case, which is the focus of our attention in this study, this formula explicitly reduces to

$$\begin{aligned} P(G_1|D) &= \frac{P(D|G_1) * P(G_1)}{P(D|G_1) * P(G_1) + P(D|G_2) * P(G_2)} && (1) \\ P(G_2|D) &= 1 - P(G_1|D) && (2) \end{aligned}$$

SPSS apparently uses the minimum χ^2 rule [8], which is the same as using the maximum posterior probability.

$$\begin{aligned} P(D|G_1) &= \exp -\frac{\chi^2}{2} \\ \text{Since } \chi^2 &= \left(\frac{x - \overline{x}}{\sigma}\right)^2 \\ \text{and } \sigma &= 1 \\ P(D|G_1) &= \exp -\frac{(x - \overline{x}_1)^2}{2} \\ P(D|G_2) &= \exp -\frac{(x - \overline{x}_2)^2}{2} \end{aligned}$$

The discriminant scores, x_i, are calculated by the SPSS/PC+ program so that $\sigma = 1$, $\overline{x}_1$ is the mean of the students who passed, and $\overline{x}_2$ is the mean of the students who failed on the canonical discriminant function. Substituting these values for $P(D|G_1)$ and $P(D|G_2)$ and where $G_1 = PASS$ and $G_2 = FAIL$ in equation 1 yields,

$$P(PASS|D) = \frac{\exp -\frac{(x-\overline{x}_1)^2}{2} * P(PASS)}{\exp -\frac{(x-\overline{x}_1)^2}{2} * P(PASS) + \exp -\frac{(x-\overline{x}_2)^2}{2} * P(FAIL)}$$

A numerical example using this formula for the calculation of the posterior probability of belonging to the PASS or FAIL group follows:

The discriminant score for the hypothetical student	$x = .1638$
Group mean of discriminant scores for PASS	$\overline{x}_1 = -.03637$
Group mean of discriminant scores for FAIL	$\overline{x}_2 = .82694$
Prior probability of PASS	$P(PASS) = .58$
Prior probability of FAIL	$P(FAIL) = .42$

$$
\begin{aligned}
P(PASS|x_i = .1638) &= \frac{\exp -\frac{(.1638-(-.03637))^2}{2} * .58}{\exp -\frac{(.1638-(-.03637))^2}{2} * .58 + \exp -\frac{(.1638-.82694)^2}{2} * .42} \\
&= \frac{.9802 * .58}{(.9802 * .58) + (.8026 * .42)} \\
&= .6278
\end{aligned}
$$

Therefore, the posterior probability of a PASS in advanced flight training given that a student achieved a discriminant function score of .1638 = .6278. From equation 2 the posterior probability of a FAIL = .3722. The probability of belonging to the PASS group, given a discriminant score of .1638, is about 63% and, therefore, this student will be classified by the program as a PASS because the PASS group has the maximum posterior probability.

The prior probabilities parameter for the PASS and FAIL groups was changed in the DA program through trial-and-error until the classification matrix yielded as close to a 50% reduction in failures as possible. The values arrived at were P(PASS)=.58 and P(FAIL)=.42, as given above. This parameter takes into account not only the prior probabilities, but also the payoff matrix for all four possible decisions. It reflects the willingness to tradeoff false rejections for a reduced number of failures and explains why the prior probabilities were not, for example, simply P(PASS)=.95 and P(FAIL)=.05.

The threshold discriminant function score, ($x_{threshold}$), above which a student will be classified as a FAIL, and below which the student will be classified as a PASS can be determined

$$
\begin{aligned}
P(PASS|x_{threshold} =?) &= .50 \\
P(PASS|x_{threshold} = .769) &= .50
\end{aligned}
$$

Any SNA who gets a discriminant function score above .769 will be predicted to fail advanced flight training; any discriminant function score below .769 will be predicted to pass.

In addition to this direct approach, logistic discrimination using the discriminant function scores can be used as well to calculate the same posterior probabilities. It is an interesting exercise to make the comparison between the parameters used above and the logistic discrimination parameters. This derivation is shown in the final section of the appendix.

Ultimately, the goal is to derive an expression for the posterior probability for the two groups as a logistic function.

$$\begin{aligned} P(G_1|D) &= \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \\ P(G_2|D) &= \frac{1}{1 + e^{\alpha + \beta x}} \end{aligned}$$

By using discriminant function scores to classify SNAs, it can be safely assumed that these scores are normally distributed with a standard deviation equal to one.

To start the derivation of the logistic function for the posterior probabilities refer back to equation 1. Dividing both the numerator and the denominator of equation 1 by $P(D|G_2) * P(G_2)$ yields,

$$P(G_1|D) = \frac{\frac{P(D|G_1)*P(G_1)}{P(D|G_2)*P(G_2)}}{\frac{P(D|G_1)*P(G_1)}{P(D|G_2)*P(G_2)} + 1}$$

Within this expression, focus attention on the likelihood ratio,

$$\frac{P(D|G_1)}{P(D|G_2)}$$

$$\begin{aligned} P(D|G_1) &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \\ P(D|G_2) &= \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \\ \text{Since } \sigma_1 &= \sigma_2 = 1 \\ \frac{P(D|G_1)}{P(D|G_2)} &= \frac{e^{-\frac{1}{2}(x-\mu_1)^2}}{e^{-\frac{1}{2}(x-\mu_2)^2}} \\ &= e^{-\frac{1}{2}(x-\mu_1)^2 + \frac{1}{2}(x-\mu_2)^2} \end{aligned}$$

Expanding both quadratic terms in the exponent:

$$\begin{aligned} (x-\mu_1)^2 &= x^2 - 2x\mu_1 + \mu_1^2 \\ (x-\mu_2)^2 &= x^2 - 2x\mu_2 + \mu_2^2 \\ -\frac{1}{2}(x-\mu_1)^2 + \frac{1}{2}(x-\mu_2)^2 &= -\frac{1}{2}x^2 + x\mu_1 - \frac{1}{2}\mu_1^2 + \frac{1}{2}x^2 - x\mu_2 + \frac{1}{2}\mu_2^2 \end{aligned}$$

$$
\begin{aligned}
&= x\mu_1 - x\mu_2 - \frac{1}{2}(\mu_1^2 - \mu_2^2) \\
\mu_1^2 - \mu_2^2 &= (\mu_1 - \mu_2)(\mu_1 + \mu_2) \\
&= x(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 - \mu_2)(\mu_1 + \mu_2) \\
\text{Let } \alpha_0 &= -\frac{1}{2}(\mu_1 - \mu_2)(\mu_1 + \mu_2) \\
\beta &= \mu_1 - \mu_2 \\
e^{-\frac{1}{2}(x-\mu_1)^2 + \frac{1}{2}(x-\mu_2)^2} &= e^{\alpha_0 + \beta x} \\
P(G_1|D) &= \frac{e^{\alpha_0+\beta x} * \frac{P(G_1)}{P(G_2)}}{e^{\alpha_0+\beta x} * \frac{P(G_1)}{P(G_2)} + 1}
\end{aligned}
$$

The posterior probability is now very close to the required form. The final step is to transform the expression

$$e^{\alpha_0 + \beta x} * P(G_1)/P(G_2)$$

The general approach is to take inverse operations in succession and form a new constant. Taking the natural log (ln) and exponentiating are two such inverse operations. The expression

$$e^{\alpha_0 + \beta x} * P(G_1)/P(G_2)$$

is rewritten as

$$e^{x} * y_1/y_2$$

Step	Operation	Result
1	Take natural log	$\ln(e^{x} * y_1/y_2) = x + \ln(y_1/y_2)$
2	Form exponential	$e^{x + \ln(y_1/y_2)} = e^{\alpha_0 + \beta x + \ln[(P(G_1)/P(G_2)]}$
3	Form new constant	$\alpha_1 = \alpha_0 + \ln[P(G_1)/P(G_2)]$
4	Posterior Probability	$P(G_1 \mid D) = \dfrac{e^{\alpha_1 + \beta x}}{1 + e^{\alpha_1 + \beta x}}$

It is quite easy at this point to show that

$$P(G_2|D) = 1 - P(G_1|D)$$

$$
\begin{aligned}
&= \frac{1+e^{\alpha_1+\beta x}}{1+e^{\alpha_1+\beta x}} - \frac{e^{\alpha_1+\beta x}}{1+e^{\alpha_1+\beta x}} \\
&= \frac{1}{1+e^{\alpha_1+\beta x}}
\end{aligned}
$$

The following numerical example shows that this derivation provides the same posterior probability for a given discriminant score, $P(PASS|x = .1638)$, as calculated before:

$$
\begin{aligned}
\alpha_1 &= \alpha_0 + \ln[P(PASS)/P(FAIL)] \\
\alpha_0 &= -\frac{1}{2}(\mu_{PASS} - \mu_{FAIL})(\mu_{PASS} + \mu_{FAIL}) \\
&= -\frac{1}{2}(-.03637 - .82694)(-.03637 + .82694) \\
&= .3413 \\
\alpha_1 &= .3413 + \ln[.58/.42] \\
&= .6640 \\
\beta &= \mu_{PASS} - \mu_{FAIL} \\
&= -.03637 - .82694 \\
&= -.8633 \\
\alpha_1 + \beta x &= .6640 - .8633(.1638) \\
&= .5226 \\
P(PASS|x = .1638) &= \frac{e^{\alpha_1+\beta x}}{1+e^{\alpha_1+\beta x}} \\
&= \frac{e^{.5226}}{1+e^{.5226}} \\
&= .6278
\end{aligned}
$$

APPENDIX C

The calculations in this appendix show how the prior and posterior probability density functions presented in Fig. 3 were derived. The numerical details of the Bayesian confidence interval calculations are also presented.

The derivation of results presented in the main section of the report follows a standard Bayesian treatment of a binomial distribution with a continuous parameter [6]. A beta distribution was chosen as a natural conjugate prior for θ, the success rate parameter. A beta distribution is characterized by two parameters, α and β, both greater than zero. The functional form for the prior distribution is

$$g(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

The prior distribution, $g(\theta)$, chosen for this analysis had $\alpha = 95$ and $\beta = 5$. The expectation of θ is

$$E(\theta) = \frac{\alpha}{\alpha+\beta} = .95$$

The variance of θ is defined as

$$var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = .0005$$

The standard deviation of θ is, therefore, .0217. These values of α and β were selected to represent the state of knowledge for success rate in advanced flight training based on historical records.

The beta function, $B(\alpha,\beta)$, is defined as

$$B(\alpha,\beta) \equiv \frac{(\alpha-1)!\,(\beta-1)!}{(\alpha+\beta-1)!}$$

The ordinate of $g(\theta)$ can, therefore, be calculated by

$$\begin{aligned}
g(\theta) &= k_1\theta^{k_2}(1-\theta)^{k_3} \\
k_1 &= \frac{(\alpha+\beta-1)!}{(\alpha-1)!\,(\beta-1)!} \\
k_2 &= \alpha-1 \\
k_3 &= \beta-1
\end{aligned}$$

A numerical example of calculating the ordinate for the prior density for a specific value of θ, for example, $\theta = .92$, would be

$$
\begin{aligned}
g(\theta = .92) &= k_1\, .92^{94}\,(1 - .92)^4 \\
k_1 &= \frac{99!}{94!\,4!} \\
&= 357,615,720 \\
&= 357,615,720 * .92^{94} * .08^4 \\
&= 5.7787
\end{aligned}
$$

The ordinate for $\theta = .92$ shown in Fig. 3 is 5.78.

The posterior density function, that is, the density function for θ after considering additional data, because of the convenient mathematical properties of the natural conjugate prior, is

$$h(\theta|y) = \frac{1}{B(y+\alpha, n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

This is seen to be of the same functional form as for the prior density function, but with the number of successes and the number of failures in the new data included. The number of successes in the new information is y, and the number of failures in the new information is $n - y$.

For a numerical example, consider the construction of the posterior probability density function, $h(\theta|y)$, for the present selection system. The ordinate of the posterior density function at $\theta = .955$ is,

Total sample size	$n = 836$
Number of PASSes	$y = 788$
Number of FAILures	$n - y = 48$
Prior density	$\alpha = 95$
Prior density	$\beta = 5$
Posterior density	$y + \alpha - 1 = 882$
Posterior density	$n - y + \beta - 1 = 52$

$$
\begin{aligned}
h(\theta = .955|y = 788) &= \frac{1}{B(883, 53)}\theta^{882}(1-\theta)^{52} \\
&= 7.8339 * 10^{88} * .955^{882} * .045^{52} \\
&= 16.7495
\end{aligned}
$$

In the second situation, the PPDF was constructed for θ when we used information from the ADHT dual task and certain biographical data. To calculate the ordinate for this PPDF at the same theoretical pass rate, $h(\theta = .955|y = 339)$,

Total sample size	$n = 348$
Number of PASSes	$y = 339$
Number of FAILures	$n - y = 9$
Prior density	$\alpha = 95$
Prior density	$\beta = 5$
Posterior density	$y + \alpha - 1 = 433$
Posterior density	$n - y + \beta - 1 = 13$

$$\begin{aligned} h(\theta = .955|y = 339) &= \frac{1}{B(434, 14)}\theta^{433}(1-\theta)^{13} \\ &= 1.6623 * 10^{27} * .955^{433} * .045^{13} \\ &= 11.3222 \end{aligned}$$

In this final section of appendix C we show how the confidence intervals were calculated. One of the advantages of the Bayesian approach is the direct meaning of the concept of a confidence interval. It is simply the desired area of the posterior probability density function, usually calculated in equal increments starting from the mode of the PPDF. The area of the PPDF for a 95% Bayesian confidence interval is represented by

$$\int_l^u \frac{1}{B(y+\alpha, n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}\, d\theta = .95$$

where l = lower limit of θ and u = upper limit of θ such that the value of the integral equals the required confidence interval, namely .95. The Mathcad software package [9] was used to vary the lower and upper limits until the desired value of the integral was reached. The lower and upper limits were systematically moved in equal increments from the mode of the particular PPDF.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1992	Interim

4. TITLE AND SUBTITLE
Performance-Based Testing and Success in Naval Advanced Flight Training

5. FUNDING NUMBERS
63706N
M0096.002-M00960.01

6. AUTHOR(S)
David J. Blower

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Aerospace Medical Research Laboratory
Naval Air Station, Bldg. 1953
Pensacola, FL 32508-5700

8. PERFORMING ORGANIZATION REPORT NUMBER
NAMRL-1378

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Center
National Naval Medical Center, Bldg. 1
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Roughly 5% of student naval aviators fail the advanced phase of flight training. At this stage of training, the Navy has spent between $300,000 and $1,000,000 per student. Any reduction in this attrition rate through prior screening would be of great economic benefit to the Navy. Computer-based performance tests developed at the Naval Aerospace Medical Research Laboratory were assessed to determine whether they could augment the present medical screening standards and thereby help identify potential failures in advanced flight training. A weak statistical relationship exists between a dual-task performance test, accession source, college major, an aptitude test, and success in advanced flight training. Discriminant analysis was employed to find a linear composite score of these variables that could be used to classify a student as a probable pass or fail in advanced flight training. For example, the model presented in this report reduced failures by 50% at the cost of rejecting roughly 20% of those students who eventually passed. A Bayesian analysis of the success rate parameter showed that this particular model significantly improved the present selection system.

14. SUBJECT TERMS
Pilot selection, Performance-based testing, Discriminant analysis, Bayesian analysis of success rate

15. NUMBER OF PAGES
35

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR